

# A compound linear discriminant method for small-magnitude seismic events and its application to the North Korea seismic event of October 9, 2006

Il-Young Che, Myung-Soon Jun, and Jeong-Soo Jeon

Earthquake Research Center, Korea Institute of Geoscience and Mineral Resources, Daejeon, Korea

(Received July 30, 2007; Revised August 27, 2007; Accepted September 4, 2007; Online published October 2, 2007)

On October 9, 2006, a seismic event ( $m_b$  3.9) occurred in North Korea, which was later announced as a nuclear test. We studied the seismic event from the viewpoint of seismological discrimination without *a priori* information of the source. This paper presents a compound discriminant method that enhances the ability to distinguish between earthquakes and explosions on the Korean Peninsula. The method includes four single discriminants based on seismic spectral amplitude ratios. These discriminants are used as single variables in a multivariate statistical analysis to derive a compound linear discriminant function that optimally separates the two populations. Short-period vertical component seismograms recorded by seismo-acoustic array (CHNAR) are used for the method. Combining the mutually complementary discrimination abilities of the single methods improves the discriminant power and lowers the probability of misclassification to 1.7% for the data sets. Using this compound method, we were able to identify the North Korea seismic event as explosion-induced. This discrimination technique can be applied automatically to new observations and extended to other seismic arrays or networks. The combination of these seismic discriminants with infrasound observations will increase our ability to detect surface explosions on the Korean Peninsula.

**Key words:** Linear discriminant method, small-magnitude seismic events, seismic spectral ratio.

## 1. Introduction

The Korea Institute of Geoscience and Mineral Resources (KIGAM), the Korean National Data Center (NDC) under the Comprehensive Nuclear Test Ban Treaty Organization (CTBTO), operates a high-quality seismic network consisting of 32 stations. In combination with the Korea Earthquake Monitoring System (KEMS), it automatically detects and analyzes seismic events and compiles daily reviews since 1998. Although seismic activity tends to be weak or moderate in and around the Korean Peninsula, KEMS has analyzed more than 3,000 seismic events annually in this region. Most events are small in magnitude ( $< 4 M_L$ ), and many are assumed to be anthropogenic in origin. Explosion contamination is a major cause of error when estimating regional seismicity. Therefore, the identification of explosion-generated seismic events is necessary to construct an accurate database of natural earthquakes as well as to identify suspicious seismic events, such as those linked with clandestine nuclear testing in violation of the CTBT.

To resolve this issue, KIGAM has been operating seismo-acoustic arrays in an attempt to identify surface explosions. Infrasound monitoring is also one of the standard technologies adopted as part of the International Monitoring System of the CTBT. Seismo-acoustic analysis, which associates seismic and infrasonic signals to those generated by surface explosions, has been used successfully to identify explo-

sions on the Korean Peninsula (Che *et al.*, 2002). However, not all explosions have been distinguished by this analysis method. Seismo-acoustic analysis has identified about 10% of the total number of seismic events that have occurred on the Korean Peninsula as surface explosions; however, many unidentified seismograms have explosion characteristics. Thus, another discriminant method is needed that can directly use the seismograms, thereby supplementing seismo-acoustic analysis and increasing the power of discrimination.

For this purpose, we developed a compound discriminant method involving seismic spectral ratio methods, which have been widely used for identifying seismic sources in various regions (Koch and Fäh, 2002; Rodgers and Walter, 2002). Regional seismic phases Pg, Lg, and Rg were used to measure phases and cross-frequency ratios in various frequency-band combinations. Multivariate statistical analysis using the spectral ratios was performed to estimate a linear discriminant function (LDF) and the discriminant potential for each spectral ratio method. Since it is desirable to fuse various methods to resolve the limited discriminant power of any single method, we included a compound linear discriminant function (CLDF), which used the single methods as variables in a multivariate analysis. This method decreases the number of misclassified events and improves the discriminant potential. Two training datasets representing earthquakes and surface explosions that occurred on the Korean Peninsula were constructed and used in deriving the discriminant functions. The dataset values were estimated based on seismic events recorded by the Cheolwon seismo-acoustic array (CHNAR) located in the central part of the

Korean Peninsula. As a test of the method, we applied it to the North Korea seismic event that occurred on October 9, 2006, and to known earthquakes and surface explosion.

## 2. Discriminant Method

Focal mechanism and source-time function analyses may explain the generation of seismic waves more quantitatively and can be used as discriminant methods. However, these methods require high-quality seismic data and well-defined responses of media in specific regions. Therefore, the application of these methods is restricted in the Korean Peninsula because many seismic events have small magnitudes and observations are by local seismic stations. Thus, we adopted the widely applied seismic spectral ratio method since it is easy to apply to a large dataset, and its statistical techniques produce effective discrimination, even though the results depend on the region studied.

Here, we use the following four single methods to discriminate small-magnitude seismic events: the Pg/Lg,  $Lg_{low}/Lg_{high}$ ,  $Pg_{low}/Pg_{high}$ , and Rg/Lg methods. Pg and Lg are the dominant seismic phases at regional distances. The Rg phase is also detected from near-surface seismic sources. Pg/Lg and Rg/Lg represent the ratios of the phase spectral amplitude, and  $Lg_{low}/Lg_{high}$  and  $Pg_{low}/Pg_{high}$  are cross-frequency amplitude ratios in the frequency range of 1 to 17 Hz. Numerators in the cross-frequency ratio method indicate spectral amplitudes in the low-frequency band, and denominators denote the high-frequency band. Time windows embracing the Pg, Lg, and Rg phases were set for the seismograms at group velocities around 5.7, 3.3, and 2.7 km/s, respectively. The Lg and Rg windows centered at the group velocities are truncated at  $\pm 1.96\sigma$  ( $\approx 95\%$ ), where standard deviation  $\sigma_{ref}$  is 2.0 s for a reference distance ( $\Delta_{ref} = 100$  km) and scaled by  $\Delta/\Delta_{ref}$  for different distances. The Pg window is scaled to the Lg window using an average propagation velocity ratio  $V_s/V_p$  ( $\approx 1/\sqrt{3}$ ) (Kim *et al.*, 1997). Since the CHNAR array (discussed later) consists of four seismic stations, the array-averaged amplitude spectrum generally produces smoother spectra for each phase; this step reduces errors caused by site effects and background noise at individual stations (Koch, 2002). The logarithms of the amplitude spectral ratios were then corrected using an attenuation model for Korea (Chung *et al.*, 2006) to remove propagation-path effects.

A multivariate discriminant analysis was then performed on these spectral ratios to derive the LDF that optimally separated the established training datasets using available multivariate information on the populations. The LDF consisted of weighted independent variables, defined as estimates of the two mean vectors of the earthquake and explosion populations. To determine the optimal functions, we used several different frequency combinations (band width = 2 Hz) in the possible frequency range. The discriminant power of each method was then estimated by the distance between the means of the two populations (Mahalanobis distance) and the probability of misclassification based on the mean distance (Kim *et al.*, 1997). The derived LDF provides an estimated discriminant score for a new observation and determines the population to which the new observation belongs. In our case, a positive score indicated the explosion

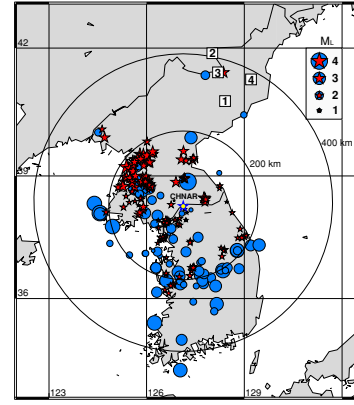


Fig. 1. Training datasets, including 82 earthquakes (solid circles) and 181 explosions (solid stars), and the location of CHNAR. The four numbered squares are additional known seismic events that were used to test the compound discriminant method. Number 4 indicates the location of the North Korea seismic event that occurred on October 9, 2006.

group and a negative score the earthquake group.

## 3. Earthquake and Explosion Training Datasets

The discriminant methods were developed using seismic events recorded by CHNAR from four GS-13 short-period vertical seismometers and 11 infrasonic sensors within a 1-km aperture. To derive reliable discriminant functions and conduct effective discriminant analyses, we constructed training datasets representing natural earthquakes and surface explosions. One problem in constructing a representative training dataset can be insufficient information for each seismic event, especially for small-magnitude events. For the explosion training dataset, we selected seismic events that had been recorded and identified as explosions in the seismo-acoustic database. For the natural earthquake training dataset, we carefully selected seismic events that took place around midnight and were recorded in the KEMS database. We considered regional coverage, various magnitudes, and data quality for each event selected. In total, we selected 82 earthquakes and 181 explosions for the training datasets; all these events took place within 400 km of CHNAR between 2001 and 2006. Figure 1 shows the distributions of the earthquake and explosion populations and the location of CHNAR. Solid circles indicate representative earthquakes, and solid stars denote explosions. Most of the explosions are concentrated in several epicentral regions. Some epicenters in South Korea were confirmed by field survey as a rock quarry, an open-pit limestone mine and a construction site. The four numbered squares show the locations of the additional seismic events in the NE of the Korean Peninsula used to test the compound discriminant function. Number 4 marks the seismic event that occurred on October 9, 2006 in North Korea.

## 4. Compound Discriminant Analysis

Four discriminant methods were separately developed based on the earthquake and explosion training datasets. Each LDF was estimated, and its power of discrimination and the probability of misclassification were calculated. For each of the single methods, the LDF could be formulated

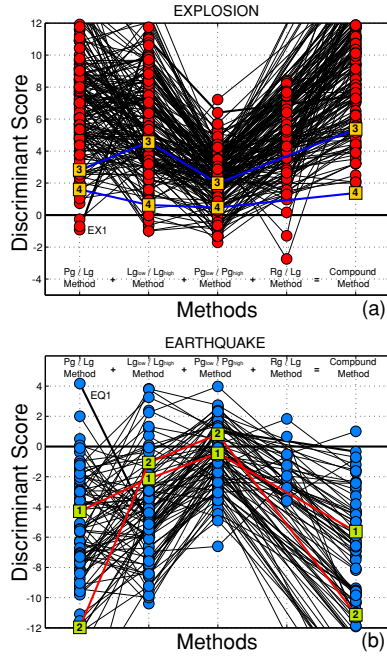


Fig. 2. (a) Discriminant scores for explosion (a) and earthquake populations (b) calculated using the four discriminant methods and the compound method (right columns).

with several weighted variables of different frequency-band combinations. To simplify the function to include only variables making significant contributions, we removed variables contributing relatively less by backward elimination, without losses to the power of discrimination (Koch and Fäh, 2002).

Of the four single methods, the Pg/Lg method was the most promising as it was highly weighted toward high-frequency bands (9–13 Hz).  $Lg_{low}/Lg_{high}$  was also useful for discrimination because the spectral contents of the explosion population were concentrated at lower frequencies (lower source corner frequency) than the spectral contents of the earthquakes. However,  $Pg_{low}/Pg_{high}$  did not show significant discriminant power. Compared to  $Lg_{low}/Lg_{high}$ , its relatively low discriminant power might have been due to the small amplitude in the Pg phase, indicating that the Pg phase has insufficient information for discrimination compared to the dominant Lg phase. This was particularly evident in the earthquake population, as more misclassification resulted at smaller Pg amplitudes than was the case for explosions. In this study, we applied the Rg/Lg method (0.5–1.5 Hz band) only for events within 100 km of CHNAR since the Rg phase was not discernible for events more than 100 km away due to strong attenuation. The resulting misclassification probabilities for the Pg/Lg,  $Lg_{low}/Lg_{high}$ ,  $Pg_{low}/Pg_{high}$ , and Rg/Lg methods were estimated as 3.6, 5.8, 18.6, and 11.2%, respectively. Although the Pg/Lg method produced the best results, it misclassified three explosions and four earthquakes. It is critical to correct such misclassifications and lower their probability so that a suspicious seismic event will not be misinterpreted.

Figure 2 shows the changes in the discriminant scores for all populations using the four methods of discrimination (left four columns). For example, the explosion sam-

ple EX1 shown in Fig. 2(a) was misclassified by the Pg/Lg method. However, this event was correctly classified by the other methods (black solid line). Other explosion samples that were misclassified or had ambiguous discriminant scores using a single method were correctly classified by the other methods. The earthquake population shown in Fig. 2(b) also showed similar results. A few misclassified or ambiguously classified samples might have resulted from factors such as the characteristics of the seismic source, propagation effect, and the signal-to-noise ratio. However, as shown in Fig. 2, even though a few samples were not correctly discriminated by an individual method, those events were correctly discriminated by other methods. This result suggests that discrimination can be improved by applying the four discriminant methods simultaneously. Accordingly, we devised a compound discriminant function based on the mutually complementary discriminant abilities of the single methods. The compound method is also a LDF, which directly uses the four methods as single variables. The values of the variables are the discriminant scores obtained from each LDF, instead of the spectral ratios used in the previously discussed single discriminant methods. When the Rg/Lg method was fused into the CLDF, as mentioned above, samples with an epicentral distance larger than 100 km were not considered. In practice, discriminant scores of zero were assigned as the Rg/Lg variable for long-distance (>100 km) events since a zero score means that the event does not belong to any two groups and would not affect the discriminant function. The CLDF obtained is  $CLDF = 0.84 Pg/Lg + 0.70 Lg_{low}/Lg_{high} + 0.15 Pg_{low}/Pg_{high} - 0.15 Rg/Lg - 0.47$ . The right-hand columns in Fig. 2 show the results of the CLDF. The method successfully identified all the samples in the training sets, except for one earthquake; the probability of misclassification decreased to 1.7%. The misidentified earthquake (22 km from CHNAR) was correctly identified by the Pg/Lg and Rg/Lg method, but misidentified by two other methods. Especially, discriminant score obtained by  $Lg_{low}/Lg_{high}$  method was particularly incorrect and about two times larger than other scores, thus it caused the misidentification in CLDF. Compared with other correctly identified local earthquakes and explosions, we could interpret that the failure of  $Lg_{low}/Lg_{high}$  method was due to deficiencies in high frequencies in Lg. In the CLDF,  $Pg_{low}/Pg_{high}$  and Rg/Lg methods made insignificant contribution to the compound discrimination, but we included all methods to slightly lower the probability of misclassification.

## 5. Application to North Korean Seismic Events

Four seismic events that occurred in North Korea were used to test the compound method. One was the North Korea seismic event of October 9, 2006; the others took place near that event (see Fig. 1). Events 1 ( $M_L$  3.5, April 16, 2002, 22:52, UTC) and 2 ( $M_L$  3.8, December 16, 2004, 18:59) were interpreted as earthquakes, while number 3 ( $M_L$  2.8, July 23, 2005, 03:08) was considered to be a typical seismo-acoustic event from a surface explosion (Fig. 3). Spectral ratios were calculated following the defined frequency bands in the established CLDF. Figure 2 presents

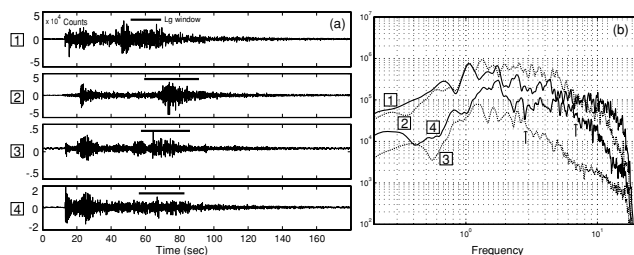


Fig. 3. Waveforms (a) and array-averaged Lg spectra (b) of the four test seismic events in Fig. 1. Arrows in (b) indicate the corner frequencies of numbers 3 and 4 test seismic events.

the estimated discriminant scores for these events. The compound method correctly distinguished the two earthquakes and one surface explosion. The North Korea event was also identified as an explosion-type seismic event rather than an earthquake.

The compound method was trained by representative populations that were highly weighted to high-frequency bands ( $> \sim 9$  Hz). The explosion population was characterized by a higher amplitude in low-frequency bands compared to the earthquake population. However, the North Korea event had a higher source corner frequency ( $\sim 7$  Hz) than the general explosion population ( $\sim 3$  Hz). Because of this spectral distribution, the compound method showed a lower discriminant power for the North Korea event than for other typical explosions, including the number 3 event. This result suggests that the North Korea event was slightly different in its source generation compared to general industrial explosions occurring in North Korea.

## 6. Conclusions

Based on the spectral characteristics of seismic phases at local and regional distances, a multivariate discriminant analysis was performed to derive a compound linear discriminant method applicable to small-magnitude seismic events occurring on the Korean Peninsula. The method was derived by adopting the mutually complementary characteristics of single seismic discriminant methods. All samples in the training datasets were correctly discriminated, except for one, and the probability of misclassification was lowered to 1.7%. Of the single methods, Pg/Lg was the most promising and was weighted toward ratios with higher frequencies (9–13 Hz).  $Lg_{\text{low}}/Lg_{\text{high}}$  also showed good discriminant power in ratios of 2–4 to 9–13 Hz.

High frequencies were deficient in the spectrum for the explosion population compared to that of the earthquake population. Therefore, Lg, which is particularly deficient in high frequencies, has worked well for discriminant applications in other regions (Murphy and Bennett, 1982; Taylor *et al.*, 1988). Gupta *et al.* (1992) and Patton and Taylor (1995) suggested that the low-frequency part of the Lg spectra from explosions is mainly due to the S-wave generated by the near-source scattering of Rg. An explanation of the low-frequency characteristics of explosions is beyond the scope of this study; however, this property can be effectively applied to distinguish explosions from natural earthquakes. The North Korea seismic event examined here differed in its spectral characteristics from the explosion population,

having significant high-frequency as well as low-frequency spectral amplitudes. Although this event was successfully identified as an explosion-type event, the resulting discriminant power was not as strong as that for other explosions because of its spectral characteristics. This suggests that the seismic source may have differed from typical explosions in North Korea (i.e., near-surface explosions generating infrasound). Thus, the effects from the depth of burial, confinement, and material properties of the seismic source region should be considered for better quantitative interpretations.

Our compound discriminant method can be applied to small-magnitude seismic events in real time in the Korea Peninsula. The method can be improved by adding other discrimination techniques, such as a method to identify ripple-firing shots, a blasting pattern that is very common in South Korea. In addition, accumulating more representative training data will improve the discriminant capability and allow for regionally specific discriminant functions for the Korean Peninsula. Although the method presented here is only applicable to CHNAR, it can be extended to other seismic arrays or local seismic networks. Developments of discrimination systems integrating the compound method and seismo-acoustic analysis should allow for the simultaneous investigation and discrimination of small-magnitude seismic events and reduce the risk of false alarms due to misclassification.

**Acknowledgments.** We would like to thank Drs. T.-S. Kang, B. Stump, and C. Hayward for helpful discussions and suggestions. We further appreciate the valuable comments of two anonymous reviewers, which helped to improve the manuscript. This work was funded by the Korea Meteorological Administration Research and Development Program under Grant CATER 2006-5103.

## References

- Che, I.-Y., M.-S. Jun, J.-S. Jeon, and K. D. Min, Analysis of local seismo-acoustic events in the Korean Peninsula, *Geophys. Res. Lett.*, **29**, 1589, doi:10.1029/2001GL014060, 2002.
- Chung, T. W., M. H. Noh, and S. Matsumoto, The attenuation structure of South Korea: a review, *J. Korean Geophys. Soc.*, **9**, 199–207, 2006.
- Gupta, I. N., W. W. Chan, and R. A. Wagner, A comparison of regional phases from underground nuclear explosions at East Kazakh and Nevada test sites, *Bull. Seismol. Soc. Am.*, **82**, 352–382, 1992.
- Kim, W. Y., V. Aharonian, A. L. Lerner-Lam, and P. G. Richards, Discrimination of earthquakes and explosions in southern Russia using regional high-frequency three-component data from the IRIS/JSP Caucasus Network, *Bull. Seismol. Soc. Am.*, **87**, 569–588, 1997.
- Koch, K., Seismic event identification of earthquakes and explosions in Germany using spectral Lg ratios, *Pure Appl. Geophys.*, **159**, 759–778, 2002.
- Koch, K. and D. Fäh, Identification of earthquakes and explosions using amplitude ratios: the Vogtland area revisited, *Pure Appl. Geophys.*, **159**, 735–757, 2002.
- Murphy, J. R. and T. J. Bennett, A discrimination analysis of short-period regional seismic data recorded at Tonto Forest Observatory, *Bull. Seismol. Soc. Am.*, **72**, 1351–1366, 1982.
- Patton, H. J. and S. R. Taylor, Analysis of spectral ratios from NTS explosions: implications for the source mechanisms of spall and the generation of Lg waves, *Bull. Seismol. Soc. Am.*, **85**, 220–236, 1995.
- Rodgers, A. J. and W. R. Walter, Seismic discrimination of the May 11, 1998 Indian nuclear test with short-period regional data from Station NIL (Nilore, Pakistan), *Pure Appl. Geophys.*, **159**, 679–700, 2002.
- Taylor, S. R., N. W. Sherman, and M. D. Denny, Spectral discrimination between NTS explosions and western United States earthquakes at regional distances, *Bull. Seismol. Soc. Am.*, **78**, 1563–1579, 1988.